PATENT ABSTRACTS OF JAPAN

(11) Publication number : 11-282837
(43) Date of publication of application : 15.10.1999

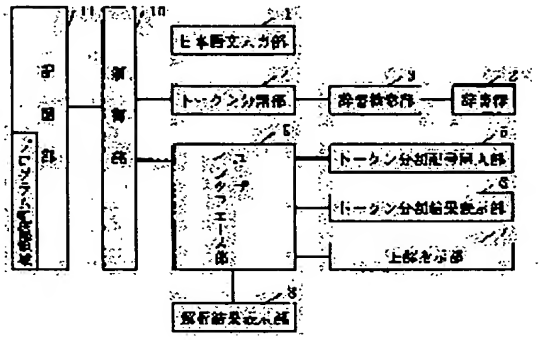
(51) Int.Cl. G06F 17/27

(21) Application number : 10-079010
(22) Date of filing : 26.03.1998

(71) Applicant : MATSUSHITA ELECTRIC IND CO LTD
(72) Inventor : KINOSHITA HITOMI

(54) JAPANESE MORPHEME ANALYSIS DEVICE AND METHOD AND RECORDING MEDIUM

(57) Abstract:
PROBLEM TO BE SOLVED: To provide a Japanese morpheme analysis device which can obtain a correct interpretation via the instruction by a user even when plural interpretations are effective.
SOLUTION: This analysis device includes a Japanese sentence input part 1 which inputs Japanese sentences as character strings, a dictionary group 2 which stores Japanese words, the information on parts of speech of the Japanese words, the vocabulary information necessary for analysis of morphemes, etc., a dictionary retrieval part 3 which retrieves the group 2 with a Japanese word used as a key, a token dividing symbol insertion part 5 which indicates the token breaks of Japanese sentences, a token dividing part 4 which divides a character string inputted via the part 1 into tokens by referring to the dictionary information and the token breaks indicated via the part 5 and a control part 10 which totally controls the analysis device.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998, 2003 Japan Patent Office

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（A）**　(11)特許出願公開番号

特開平11－282837

(43)公開日　平成11年(1999)10月15日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ０６Ｆ　17/27		Ｇ０６Ｆ　15/38	Ｅ

審査請求　未請求　請求項の数５　ＯＬ　（全 10 頁）

(21)出願番号　特願平10－79010

(22)出願日　平成10年(1998)３月26日

(71)出願人　000005821
松下電器産業株式会社
大阪府門真市大字門真1006番地

(72)発明者　木下　ひとみ
大阪府門真市大字門真1006番地　松下電器産業株式会社内

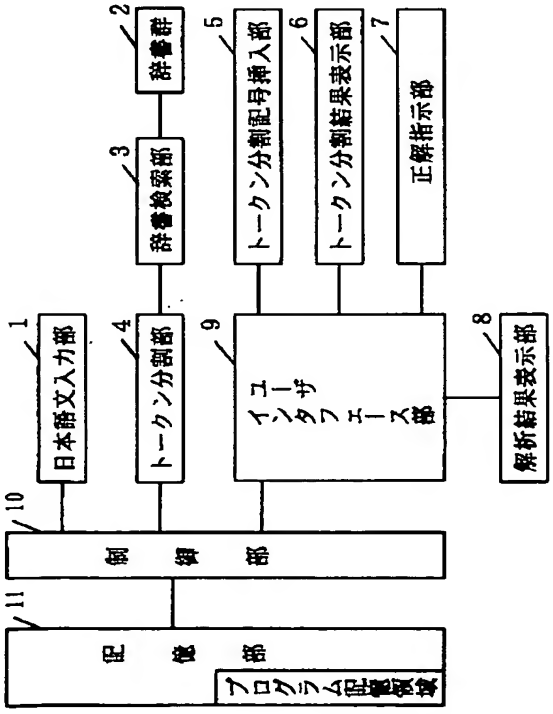
(74)代理人　弁理士　滝本　智之　（外１名）

(54)【発明の名称】　日本語形態素解析装置、日本語形態素解析方法および記録媒体

(57)【要約】

【課題】　複数の解釈を有する場合であっても、ユーザ指示により、正しい解釈を得ることができる日本語形態素解析装置を提供すること。

【解決手段】　日本語文を文字列として入力する日本語文入力部１と、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群２と、日本語単語をキーとして辞書群２を検索する辞書検索部３と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部５と、日本語文入力部から入力された文字列を辞書情報とトークン分割記号挿入部５で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部４と、全体を制御する制御部１０とを有する。



1

【特許請求の範囲】

【請求項１】日本語文を文字列として入力する日本語文入力部と、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして前記辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、前記日本語文入力部から入力された文字列を前記辞書情報と前記トークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することを特徴とする日本語形態素解析装置。

【請求項２】日本語文を文字列として入力する日本語文入力部と、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして前記辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、前記日本語文入力部から入力された文字列を最初は前記辞書情報のみを用いてトークンに分割し、前記辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には前記曖昧性が生じた部分に対して前記辞書情報と前記トークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することを特徴とする日本語形態素解析装置。

【請求項３】日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、日本語文のトークンの切れ目を指示する指示ステップと、前記入力ステップで入力された文字列を前記辞書情報と前記指示ステップで指示したトークンの切れ目とを参照してトークンに分割する分割ステップとを有することを特徴とする日本語形態素解析方法。

【請求項４】日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、前記入力ステップで入力された文字列を最初は前記辞書情報のみを用いてトークンに分割し、前記辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には日本語文のトークンの切れ目を指示すると共に前記曖昧性が生じた部分に対して前記辞書情報と前記指示されたトークンの切れ目とを参照してトークンに分割する分割ステップとを有することを特徴とする日本語形態素解析方法。

【請求項５】請求項３又は４に記載された日本語形態素解析方法を実行させるためのプログラムを記録した記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、文字列として入力した日本語文の形態情報を出力する日本語形態素解析装置、日本語形態素解析方法およびその方法を実行させるためのプログラムを記録する記録媒体に関する。

2

【０００２】

【従来の技術】ワープロのかな漢字変換や機械翻訳などの日本語処理においては、まず形態素解析を行う必要がある。形態素解析では普通、単語をキーとしてその語彙情報を記憶した辞書を検索しながら、文字列を形態素（意味を持つ最小の単位であり、この最小の単位を以下「トークン」と呼ぶ）に分割し（トークン分割）、このトークンに品詞、活用などの形態情報を付加する。形態素解析には、文節数最小法、左最長一致法、コスト最小法等の手法があり、これらの手法を用いて曖昧性を解消している。しかし、どの手法も完全ではなく、誤解釈を導くことがある。たとえば、「怪我しないように気を付ける」という日本語文の「怪我しないように」の部分をトークンに分割する場合を考えてみると、“怪我し／ないよう／に”という第１のトークン分割文、“怪我し／ない／よう／に”という第２のトークン分割文、“怪我／し／ないよう／に”という第３のトークン分割文、“怪我／し／ない／よう／に”という第４のトークン分割文などのように複数の解釈が存在することが分かる。尚、このトークン分割で参照した辞書には、サ行変格活用動詞「怪我する」、「する」、普通名詞「怪我」、「ないよう(内容)」、助動詞「ない」、形式名詞「よう」、格助詞「に」等が登録されているものとする。

【０００３】例文の場合、正解は第２の分割文もしくは第４の分割文であるが、前述の３つの手法で評価してみると、文節数最小法では“怪我し／ないよう／に”という第２の分割文となり、左最長一致法では“怪我し／ないよう／に”という第１の分割文、接続コスト最小法ではコストの付け方によってどの分割文かが定まることとなり、誤解釈を出力することになる。また、これらのどの手法を採っても、経験則に依る所が大きく、多種多様な状況を表現し得る自然言語を処理する場合、誤解釈を導くことは避けられない。

【０００４】

【発明が解決しようとする課題】このように、従来の日本語形態素解析方法では、前後の文脈情報を用いない限り正解を導き出すのは難しいという問題点を有している。また、文脈処理の技術は、実用化レベルに達していないのが現状である。

【０００５】この日本語形態素解析装置、日本語形態素解析方法および記録媒体では、複数の解釈（トークン分割結果）を有する場合であっても、ユーザ指示により、正しい解釈を得ることができることが要求されている。

【０００６】本発明は、複数の解釈を有する場合であっても、ユーザ指示により、正しい解釈を得ることができる日本語形態素解析装置、および、複数の解釈を有する場合であっても、ユーザ指示により、正しい解釈が得られる日本語形態素解析方法、ならびに、その日本語形態

素解析方法が実現される記録媒体を提供することを目的とする。

【０００７】

【課題を解決するための手段】この課題を解決するために本発明の日本語形態素解析装置は、日本語文を文字列として入力する日本語文入力部と、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、日本語文入力部から入力された文字列を辞書情報とトークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有する構成を備えている。

【０００８】これにより、複数の解釈を有する場合であっても、ユーザ指示により、正しい解釈を得ることができる日本語形態素解析装置が得られる。

【０００９】この課題を解決するための本発明の日本語形態素解析方法は、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、日本語文のトークンの切れ目を指示する指示ステップと、入力ステップで入力された文字列を辞書情報と指示ステップで指示したトークンの切れ目とを参照してトークンに分割する分割ステップとを有する構成を備えている。

【００１０】これにより、複数の解釈を有する場合であっても、ユーザ指示により、正しい解釈が得られる日本語形態素解析方法が得られる。

【００１１】この課題を解決するための記録媒体は、上記日本語形態素解析方法を実行させるためのプログラムを記録した構成を備えている。

【００１２】これにより、上記日本語形態素解析方法が実現される記録媒体が得られる。

【００１３】

【発明の実施の形態】本発明の請求項１に記載の発明は、日本語文を文字列として入力する日本語文入力部と、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、日本語文入力部から入力された文字列を辞書情報とトークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することとしたものであり、入力された日本語文の文字列が辞書情報とユーザからの指示情報との両方に基づいてトークン分割されるので、日本語文のトークン分割の曖昧性が解消されるという作用を有する。

【００１４】請求項２に記載の発明は、日本語文を文字列として入力する日本語文入力部と、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、日本語文入力部から入力された文字列を最初は辞書情報のみを用いてトークンに分割し、辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には曖昧性が生じた部分に対して辞書情報とトークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することとしたものであり、辞書情報のみを用いたトークン分割結果に曖昧性が生じた場合のみにユーザ指示が行われ、ユーザの負担が軽減されるという作用を有する。

【００１５】請求項３に記載の発明は、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、日本語文のトークンの切れ目を指示する指示ステップと、入力ステップで入力された文字列を辞書情報と指示ステップで指示したトークンの切れ目とを参照してトークンに分割する分割ステップとを有することとしたものであり、入力された日本語文の文字列が辞書情報とユーザからの指示情報との両方に基づいてトークン分割されるので、日本語文のトークン分割の曖昧性が解消されるという作用を有する。

【００１６】請求項４に記載の発明は、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、入力ステップで入力された文字列を最初は辞書情報のみを用いてトークンに分割し、辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には日本語文のトークンの切れ目を指示すると共に曖昧性が生じた部分に対して辞書情報と指示されたトークンの切れ目とを参照してトークンに分割する分割ステップとを有することとしたものであり、辞書情報のみを用いたトークン分割結果に曖昧性が生じた場合のみにユーザ指示が行われ、ユーザの負担が軽減されるという作用を有する。

【００１７】請求項５に記載の発明は、請求項３又は４に記載された日本語形態素解析方法を実行させるためのプログラムを記録することとしたものであり、記録したプログラムの実行により、請求項３又は４に記載された日本語形態素解析方法が実現されるという作用を有する。

【００１８】以下、本発明の実施の形態について、図１～図８を参照しながら説明する。

（実施の形態１）図１は、本発明の実施の形態１による日本語形態素解析装置を示すブロック図である。

【００１９】図１において、１は形態素解析処理対象の日本語文字列を入力する日本語文入力部、２は日本語単

語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶している辞書群、３は日本語単語をキーとして辞書群２を検索する辞書検索部、４は辞書検索部３の検索結果およびユーザから指示されたトークンの切れ目の情報を参照して日本語文入力部１で入力された文字列をトークンに分割するトークン分割部、５は入力文のトークンの切れ目をユーザから指示してもらうトークン分割記号挿入部、６はトークン分割部４でトークンを分割した結果、曖昧性を含むものを表示し、ユーザからの正解入力の指示を仰ぐトークン分割結果表示部、７はトークン分割結果表示部６で表示したリストの中から、正解をユーザに指示してもらう正解指示部、８は形態素解析結果をユーザに示す解析結果表示部、９はトークン分割記号挿入部５、トークン分割結果表示部６、正解指示部７、解析結果表示部８を制御するユーザ・インタフェース部、１０は日本語文入力部１、トークン分割部４、ユーザ・インタフェース部９を制御する制御部、１１は

日本語文入力部１で入力された文、辞書検索部３の検索結果、トークン分割部４で分割されたトークンデータ、トークン分割記号挿入部５でユーザより指示されたトークンの切れ目の情報および日本語形態素解析結果を記憶する記憶部である。

【００２０】ここで、トークン分割記号挿入部５におけるユーザからの指示は、トークンの切れ目を表す記号のユーザからの入力である（トークンの切れ目を表す記号として、半角の゛を用いることにする）。また、トークン分割結果表示部６においては、曖昧性を含むものは、解釈可能なトークン分割結果を示すリストとして表示される。

【００２１】（表１）に、本実施の形態で用いる辞書群２に記憶した辞書情報の一例を示す。

【００２２】

【表１】

辞書見出し	品詞	活用型	活用形
さ	動詞	サ行変格	未然形
し	動詞	サ行変格	未然形
し	動詞	サ行変格	未然形
しろ	動詞	サ行変格	命令形
する	動詞	サ行変格	終止形
する	動詞	サ行変格	連体形
すれ	動詞	サ行変格	仮定形
せよ	動詞	サ行変格	命令形
な	助動詞	形容詞	語幹
ないよう	普通名詞	なし	なし
に	格助詞	なし	なし
よう	形式名詞	なし	なし
怪我	普通名詞	なし	なし
怪我	動詞	サ行変格	語幹

【００２３】（表１）においては左から、「辞書見出し」「品詞」「活用型」「活用形」の順に示している。ここで、活用する語（動詞、形容詞、形容動詞、助動詞など）の場合は、辞書見出しとして、語幹（活用しない）部分のみを登録する。また、語幹が存在しないもの（「する」や「くる」など）は、その活用形を全て登録しておく。

【００２４】図２は本発明の実施の形態１による日本語形態素解析装置の具体的構成を示すブロック図である。

【００２５】図２において、２１はキーボードやポインティング・デバイスなどの入力装置、２２は陰極線管ディスプレイ（ＣＲＴ）や液晶ディスプレイ（ＬＣＤ）などの表示装置、２３は装置を制御する中央処理装置（ＣＰＵ）、２４はデータを一時的に記憶するランダム・アクセス・メモリ（ＲＡＭ）、２５はＣＰＵ２３が実行するプログラムを格納するリード・オンリー・メモリ（ＲＯＭ）、２６は２次記憶装置、２７は読取装置、２８はＣＤ－ＲＯＭ等の記録媒体である。

【００２６】ここで、図１の装置構成と図２の装置構成との関係を説明する。図１および図２において、日本語入力部１、トークン分割記号挿入部５、正解指示部７は入力装置２１により実現され、トークン分割結果表示部

６、解析結果表示部８は表示装置２２により、記憶部１１はＲＡＭ２４により実現される。また、辞書群２はＲＡＭ２４、ＲＯＭ２５、２次記憶装置２６のいずれかにより実現される。辞書検索部３、トークン分割部４、ユーザ・インタフェース部９、制御部１０は、ＣＰＵ２３がＲＡＭ２４およびＲＯＭ２５とデータのやり取りを行いながら、ＲＯＭ２５に記憶された各種のプログラムを実行することにより実現される。なお、本実施の形態では、ＣＰＵ２３がＲＯＭ２５に記憶されたプログラムを実行する形態をしているが、ＣＰＵ２３が実行するプログラムは、読取装置２７を用い、ＣＤ－ＲＯＭ（コンパクト・ディスク－リード・オンリー・メモリ）などの記録媒体２８に記録されたプログラムを実行する形態であっても構わない。このように構成することにより、本実施の形態は汎用コンピュータなどにおいて容易に実現が可能となる。

【００２７】以上のように構成された日本語形態素解析装置について、以下その動作を図３、図４、図６、図７に基づいて説明する。図３、図４は本実施の形態による日本語形態素解析装置の動作を示すフローチャートであり、図３は、入力文のトークンの切れ目をユーザから指示された後に、その情報を用いてトークン分割を行う処理の流れを示すものである。また、図６は本実施の形態における形態素解析結果を現すトークン・リストを示すリスト図であり、図７は本実施の形態における辞書検索結果を表すリストを示すリスト図である。なお、これらのフローチャートは、ＣＰＵ２３がＲＯＭ２５に記憶されたプログラムを実行する様子を示すものである。

【００２８】図３においてまず、日本語文入力部１から形態素解析処理対象の日本語文（文字列）を入力する（Ｓ１）。ここでは、“怪我しないように気を付ける”という文（以下、「入力文ａ」という）が入力されたものとする。また、ここで入力される文は、全て全角文字とする（すなわち、半角カタカナやAsciiコードは含まない）。

【００２９】次に、入力文ａのトークンの切れ目をユーザが指示する（Ｓ２）。ここでは、半角の＾を用いてトークンの切れ目をユーザが入力する。従来例で前述したことから、入力文ａの正しいトークンの切れ目は、“怪我し＾ない＾よう＾に＾気を付ける＾。”の第１分割文、または“怪我＾し＾ない＾よう＾に＾気を付ける＾。”の第２分割文のいずれかである。しかし、ユーザがいつも正しくトークンを認識できるとは限らない。ここでは、以下のように、敢えて一部誤った指示を行ったものとする。

【００３０】すなわち“怪我しないように＾気を付ける＾。”をユーザの指示とする。ステップ３～ステップ１５では、ステップ２でのユーザの指示および辞書情報を用いて、入力文ａのトークン分割を行う。

【００３１】まずステップ３では、分割処理中の文字位置を示す変数POSに初期値として１をセットする。次に、この変数POSと入力文ａの文字数（文の長さ）を比較し（Ｓ４）、変数POSの方が大きければトークン分割結果を表示して（Ｓ１５）、処理を終え、そうでなければステップ５以降の処理を行う。

【００３２】ステップ５では、辞書検索キーを取得し、変数keyLenに検索キーの文字数をセットする。検索キーは、変数POSが示す文字位置から最初に見付かる＾までの文字列となる。従って、「怪我しないように」が最初の検索キーで、keyLenは８となる。

【００３３】次に、ステップ５で取得した検索キーで辞書群２を検索し（Ｓ６）、検索キーが辞書に存在したか否か（そのキーで検索成功か否か）をチェックし（Ｓ７）、存在したならばステップ１３へ、存在しなければステップ８へ移る。

【００３４】ステップ８では、検索キーを図４のフローチャートで示す辞書検索ルーチンへ渡し、辞書検索処理を行う。最初の検索キー「怪我しないように」は、（表１）に示した辞書に登録されていないため、ここでの処理対象となる。

【００３５】次に、ステップ８の辞書検索の結果、トークン分割に曖昧性が生じたか否かチェックし（Ｓ９）、曖昧性があればステップ１０へ、曖昧性がなければステップ１２へ移る。

【００３６】ステップ１０では、解釈可能なトークン分割結果を全て表示する。ここでは、トークン分割結果の第１文として“怪我し／ないよう／に”、第２文として“怪我し／ない／よう／に”、第３文として“怪我／し／ないよう／に”、第４文として“怪我／し／ない／よう／に”という４種類の分割結果が表示される。

【００３７】この４種類の中から、正しいものをユーザが指示する（Ｓ１１）。次に、こうして得られた正しいトークン分割結果をトークン・リストに登録する（Ｓ１２）。

【００３８】次に、変数POSに今処理した検索キーの長さ（keyLen）を加え、ステップ４に戻る（Ｓ１４）。次のサイクル（ステップ４～ステップ１１）では、「ように」を処理することになる。つまり、ステップ５で、「ように」が検索キーとして取得される。こうして、変数POSが入力文の長さを越えるまで、処理を続けた結果、得られるトークン分割結果（トークン・リスト）が図６に示したものである。

【００３９】ステップ７から移行するステップ１３では、検索結果（品詞や活用形等のトークン情報）をトークン・リストに登録する。トークン・リストの一例を図６に示す。トークン・リストの個々のセルが、分割したトークンを表し、隣り合うトークンは双方向のポインタで連結されている。そして、そのセル内には、対応するトークンの形態情報が格納されている。

【００４０】次に、図４を用いて辞書検索処理の流れを

説明する。ここでの辞書検索では、入力文のある位置から始まる全てのトークン（辞書見出しと一致するもの）を切り出し、図７に示すようなリストとして出力する全解探索の手法を取る。実際は、この手法に加えて接続可否の検証を行ったりする必要があるが、本質的なものではないので、説明を省く。今、本ルーチンに、処理対象文字列として「怪我しないように」が渡されたものとする。

【００４１】まず、処理中の文字位置を示す変数POSおよび検索キーの長さを示す変数keyLenに初期値として１をセットし（Ｓ２１）、入力文字列の長さ（ここでは８文字）を示す変数lastPosに入力文字列の長さをセットする。次に、変数POSと変数lastPosを比較し、変数POSの方が大きければ処理を終え、そうでなければステップ２３以降の処理を行う（Ｓ２２）。

【００４２】スッテプ２３では、変数keyLenと変数lastPosを比較し、変数keyLenの方が大きければステップ２９へ、そうでなければステップ２４へ移る。

【００４３】ステップ２４では、検索キーを取得する。検索キーは、変数POSが示す文字位置から変数keyLen分の長さの文字列である（最初の検索キーはPOS(＝１）からkeyLen（＝１）文字なので「怪」）。次に、ステップ２４で取得した検索キーで辞書群２を検索する（Ｓ２５）。次いで、検索キーが辞書に存在したか否か（そのキーで検索成功か否か）をチェックし（Ｓ２６）、存在したならばステップ２７へ、存在しなければステップ２８へ移る。

【００４４】ステップ２７では、検索に成功した文字列およびその形態情報を図７のようにリスト化する。図７に示した図は、「怪我しないように」を辞書群２を用いてトークンに分割した結果を示したものである。それぞれのセルは辞書検索に成功した文字列であり、その検索結果およびトークン文字列の先頭と末尾の文字のオフセット（入力文の先頭からの文字位置）が形態情報として記憶される。更に、後接するトークンの候補へのポインタを可能な限り記憶する。例えば、いま変数POSが４で、「ない」が検索されたとする。それぞれのリストの最後のセルでそのトークン文字列の末尾の文字のオフセットがPOS－１の値と一致するセルを探し、そのセルの次に連結する。

【００４５】ステップ２８では、変数keyLenの値を１増やし、ステップ２３へ戻る。ステップ２３から移行するステップ２９では、変数POSから始まる文字列が１つでも辞書検索に成功したか否かチェックし、検索成功が１つもなければステップ３０へ、検索成功が１つでも存在すればステップ３１へ移る。

【００４６】ステップ３０では、検索成功がない場合には変数POS番目の文字を未知語として図７のリストに登録する。次に、変数POSの値を１増やし（Ｓ３１）、変数keyLenに１をセットし（Ｓ３２）、ステップ２２へ戻る。

【００４７】こうして最後まで処理した結果を図７にリストとして示す。以上のように本実施の形態によれば、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群２を用い、まず日本語文を文字列として入力し、日本語文のトークンの切れ目をユーザが指示し、入力された日本語文の文字列を辞書情報とユーザが指示したトークンの切れ目とを参照してトークンに分割するようにしたことにより、入力された日本語文の文字列が辞書情報とユーザからの指示情報との両方に基づいてトークン分割されるので、日本語文のトークン分割の曖昧性を解消することができ、複数の解釈（トークン分割結果）を有する場合であっても、ユーザ指示により、正しい解釈を得ることができる。

【００４８】（実施の形態２）本発明の実施の形態による日本語形態素解析装置は図１、図２と同様の構成であるので、その説明は省略する。本実施の形態による日本語形態素解析装置が実施の形態１と異なるところは、実施の形態１ではステップ２に示すようにまずユーザがトークンの切れ目を指示するのに対し、まず装置がトークン分割を行い、その結果トークン分割に曖昧性が生じた場合にその曖昧性が生じた箇所をユーザに示し、その中からユーザが正解を選択するという点である。

【００４９】以上のような本実施の形態について、図５を用いて説明する。図５は本実施の形態による日本語形態素解析装置の動作を示すフローチャートである。

【００５０】図５においてまず、日本語文入力部１から、形態素解析処理対象の日本語文（文字列）を入力する（Ｓ４１）。ここでも、実施の形態１と同様に、“怪我しないように気を付ける。”という文（以下、「入力文ａ」という）が入力されたものとする。また、ここで入力される文も、全て全角文字とする（すなわち、半角カタカナやAsciiコードは含まない）。次に、図４を用いて説明した辞書検索処理を行う（Ｓ４２）。辞書検索の結果、トークン分割に曖昧性が生じたか否かチェックし、曖昧性があればステップ４４へ、曖昧性がなければ、ステップ４６へ移る（Ｓ４３）。

【００５１】ステップ４４では、解釈可能なトークン分割結果を全て表示する。ここでは、トークン分割結果の第１文として“怪我し／ない／よう／に／気を付ける／。”、第２文として“怪我し／ない／よう／に／気を付ける／。”、第３文として“怪我／し／ないよう／に／気を付ける。”、第４文として“怪我／し／ない／よう／に／気を付ける／。”という４種類の分割結果が表示される。この中から、正しいものをユーザが指示する（Ｓ４５）。そして、こうして得られたトークン分割結果および個々のトークンの形態情報を表示して、処理を終える（Ｓ４６）。

【００５２】以上のように本実施の形態によれば、日本語単語、その品詞情報、形態素解析に必要な語彙情報等

の辞書情報を記憶する辞書群２を用い、日本語文を文字列として入力し、入力された日本語の文字列を最初は辞書情報のみを用いてトークンに分割し、辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には日本語文のトークンの切れ目を指示すると共に曖昧性が生じた部分に対して辞書情報と指示されたトークンの切れ目とを参照してトークンに分割するようにしたことにより、辞書情報のみを用いたトークン分割結果に曖昧性が生じた場合のみにユーザ指示が行われ、ユーザの負担を軽減することができる。

【００５３】

【発明の効果】以上のように本発明の請求項１に記載の日本語形態素解析装置によれば、日本語文を文字列として入力する日本語文入力部と、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、日本語文入力部から入力された文字列を辞書情報とトークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することにより、入力された日本語文の文字列が辞書情報とユーザからの指示情報との両方に基づいてトークン分割されるので、日本語文のトークン分割の曖昧性を解消することができ、複数の解釈（トークン分割結果）を有する場合であっても、ユーザ指示により、正しい解釈を得ることができるという有利な効果が得られる。

【００５４】請求項２に記載の発明によれば、日本語文を文字列として入力する日本語文入力部と、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、日本語文入力部から入力された文字列を最初は辞書情報のみを用いてトークンに分割し、辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には曖昧性が生じた部分に対して辞書情報とトークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することにより、辞書情報のみを用いたトークン分割結果に曖昧性が生じた場合のみにユーザ指示が行われるので、ユーザの負担を軽減することができるという有利な効果が得られる。

【００５５】本発明の請求項３に記載の日本語形態素解析方法によれば、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、日本語文のトークンの切れ目を指示する指示ステップと、入力ステップで入力された文字列を辞書情報と指示ステップで指示したトークンの切れ目とを参照してトークンに分割する分割ステップとを有することにより、入力された日本語文の文字列が辞書情報とユーザからの指示情報との両方に基づいてトークン分割されるので、日本語文のトークン分割の曖昧性を解消することができ、複数の解釈（トークン分割結果）を有する場合であっても、ユーザ指示により、正しい解釈を得ることができるという有利な効果が得られる。

【００５６】請求項４に記載の発明によれば、日本語単語，その品詞情報，形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、入力ステップで入力された文字列を最初は辞書情報のみを用いてトークンに分割し、辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には日本語文のトークンの切れ目を指示すると共に曖昧性が生じた部分に対して辞書情報と指示されたトークンの切れ目とを参照してトークンに分割する分割ステップとを有することにより、辞書情報のみを用いたトークン分割結果に曖昧性が生じた場合のみにユーザ指示が行われるので、ユーザの負担を軽減することができるという有利な効果が得られる。

【００５７】本発明の請求項５に記載の記録媒体によれば、請求項３又は４に記載された日本語形態素解析方法を実行させるためのプログラムを記録することにより、記録したプログラムを実行すれば、請求項３又は４に記載された日本語形態素解析方法を実現することができるという有利な効果が得られる。

【図面の簡単な説明】

【図１】本発明の実施の形態１による日本語形態素解析装置を示すブロック図

【図２】本発明の実施の形態１による日本語形態素解析装置の具体的構成を示すブロック図

【図３】本発明の実施の形態１による日本語形態素解析装置の動作を示すフローチャート

【図４】本発明の実施の形態１、２による日本語形態素解析装置の動作を示すフローチャート

【図５】本発明の実施の形態２による日本語形態素解析装置の動作を示すフローチャート

【図６】本発明の実施の形態１における形態素解析結果を現すトークン・リストを示すリスト図

【図７】本発明の実施の形態１、２における辞書検索結果を表すリストを示すリスト図

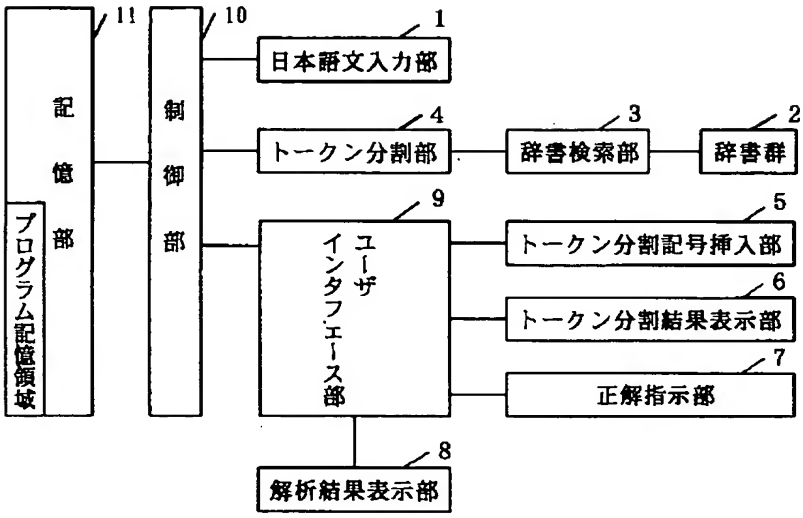
【符号の説明】

１　日本語文入力部
２　辞書群
３　辞書検索部
４　トークン分割部
５　トークン分割記号挿入部
６　トークン分割結果表示部
７　正解指示部
８　解析結果標示部

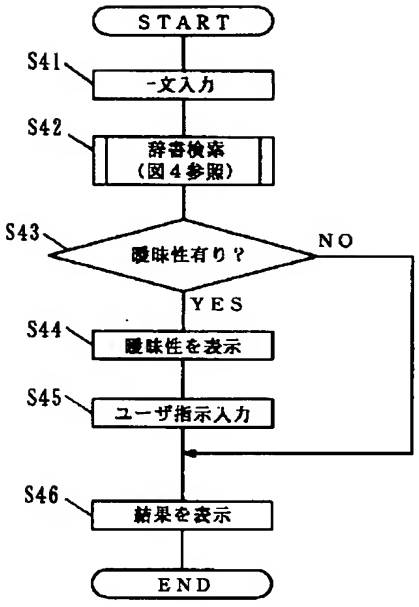
13　　14

９　ユーザ・インタフェース部
１０　制御部
１１　記憶部

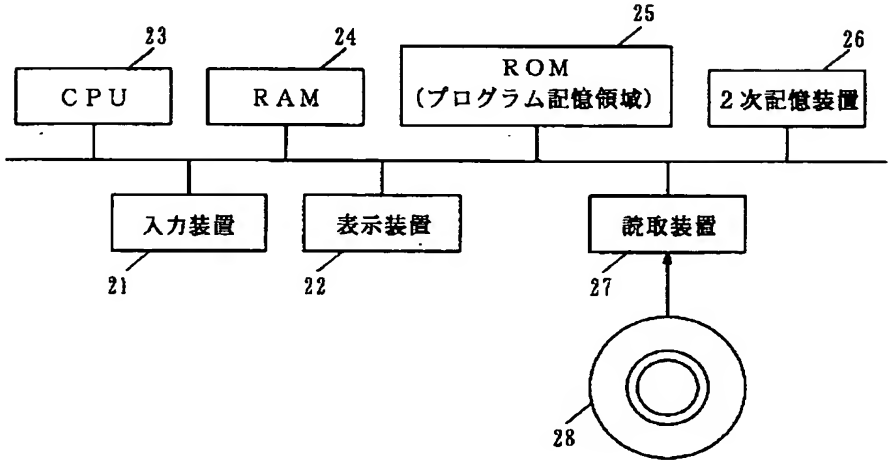
【図１】



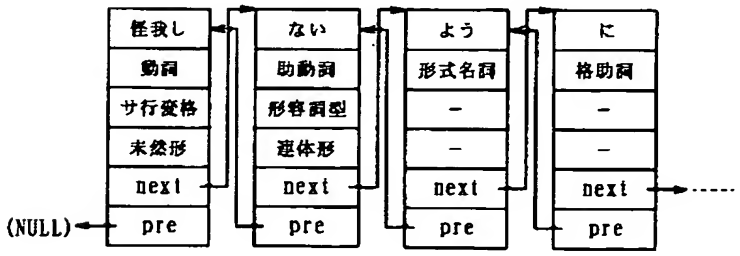
【図５】



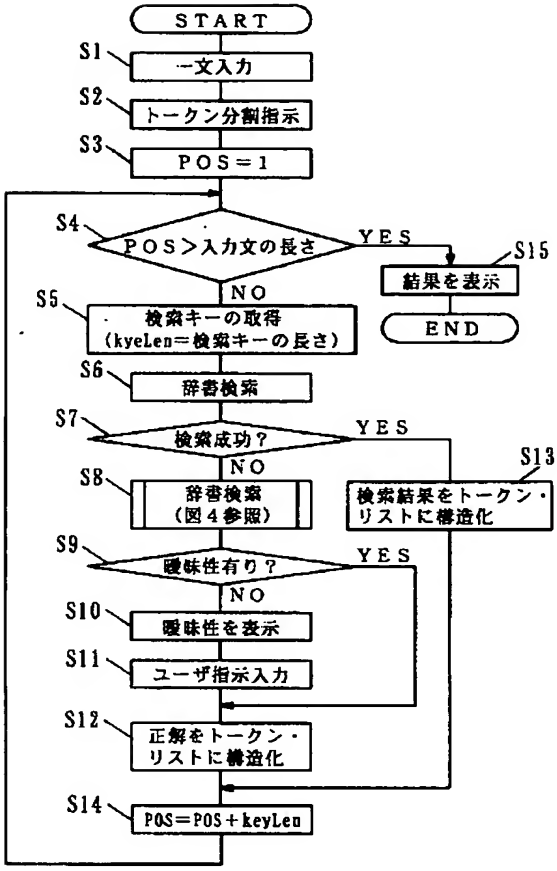
【図２】



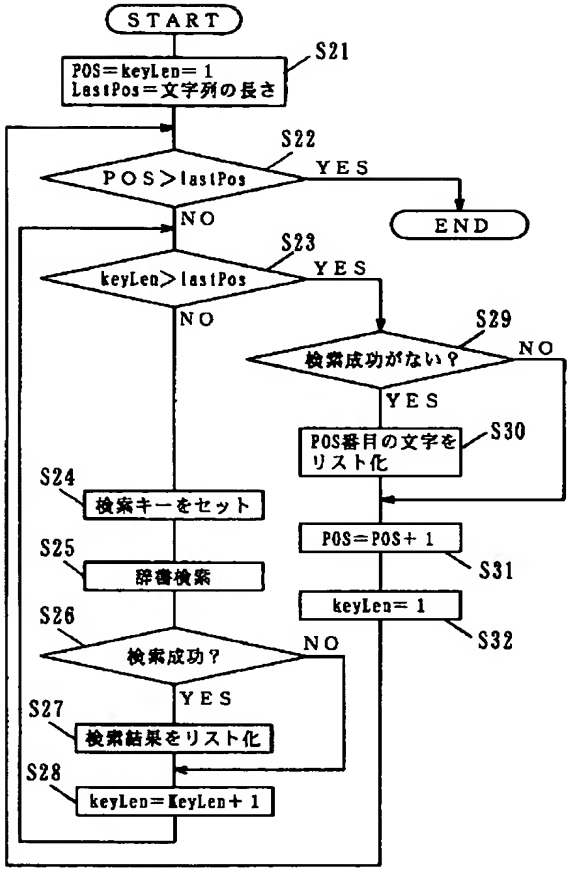
【図６】



【図３】


START
S1
一文入力
S2
トークン分割指示
S3
POS＝1
S4
POS＞入力文の長さ
YES
S15
結果を表示
END
NO
S5
検索キーの取得
(kyeLen＝検索キーの長さ)
S6
辞書検索
S7
検索成功？
YES
S13
検索結果をトークン・リストに構造化
NO
S8
辞書検索
(図４参照)
S9
曖昧性有り？
YES
NO
S10
曖昧性を表示
S11
ユーザ指示入力
S12
正解をトークン・リストに構造化
S14
POS＝POS＋keyLen

【図４】


START
S21
POS＝keyLen＝1
LastPos＝文字列の長さ
S22
POS＞lastPos
YES
END
NO
S23
keyLen＞lastPos
YES
NO
S29
検索成功がない？
NO
YES
S30
POS番目の文字をリスト化
S24
検索キーをセット
POS＝POS＋1
S31
S25
辞書検索
keyLen＝1
S32
S26
検索成功？
NO
YES
S27
検索結果をリスト化
S28
keyLen＝KeyLen＋1

【図７】